CTCs-derived xenograft development in a Triple Negative breast cancer case

Title: CTCs-derived xenograft development in a Triple Negative breast cancer case.

Tais Pereira-Veiga[1,9], Manuel Abreu[2], Diego Robledo[3], Xavier Matias-Guiu[4,5], María Santacana[4], Laura Sánchez[6], Juan Cueva[7], Patricia Palacios[7], Ihab Abdulkader[8], Rafael López-López'[1,2,5,7], Laura Muinelo-Romay[1,2,5*] and Clotilde Costa[1,5*]

1. Roche-Chus Joint Unit. University Hospital of Santiago. Travesía da Choupana s/n 15706. Santiago de Compostela, Spain.

2. Liquid Biopsy Analysis Unit, Oncomet, Health Research Institute of Santiago (IDIS), Trav. Choupana s/n 15706, Santiago de Compostela, Spain.

3. The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, EH25 9RG,UK.

4. Department of Pathology and Molecular Genetics/Oncologic Pathology Group, Arnau de Vilanova University Hospital, University of Lleida, IRB Lleida, Lleida, Spain.

5. CIBERONC, Centro de Investigación Biomédica en Red Cáncer, Madrid, Spain.

6. Department of Zoology, Genetics and Physic Antrophology. Universidade de Santiago de Compostela, Spain.

7. Department of Oncology. Complexo Hospitalario Universitario de Santiago de Compostela (SERGAS). Trav. Choupana s/n, Santiago de Compostela 15706, Spain.

8. Department of Pathology, Complexo Hospitalario Universitario de Santiago de Compostela (SERGAS), Trav. Choupana s/n, Santiago de Compostela 15706, Spain.

9. Universidade de Santiago de Compostela, Spain

*Corresponding authors: Laura.Muinelo.Romay@sergas.es (+34986955073); Clotilde.Costa.Nogueira@sergas.es (+34986955602)

Contact Information:

- Clotilde Costa: +34986955602, clotilde.costa.nogueira@sergas.es

Roche-Chus Joint Unit. University Hospital of Santiago. Travesía da Choupana s/n 15706. Santiago de Compostela, Spain.

- Laura Muinelo-Romay: +34986955073, Laura.Muinelo.Romay@sergas.es

Liquid Biopsy Analysis Unit, Oncomet, Health Research Institute of Santiago (IDIS), Trav. Choupana s/n 15706, Santiago de Compostela, Spain.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ijc.32001

Keywords: triple-negative breast cancer, Circulating Tumor Cells (CTCs), CTCs-Derived xenograft (CDX), metastasis

Abbreviations

BC, breast cancer; **CDX**, CTCs-derived xenograft; **CTCs**, Circulating Tumor cells; **ER**, estrogen receptor; **FFPE**, Formalin-fixed paraffin-embedded; **GO**, gene ontology; **HER2,** Human Epidermal growth factor-2; **M1**, metastatic site 1; **M2**, metastatic site 2; **N,** healthy or normal tissue; **PBMCs**, peripheral blood mononuclear cells; **PCA**, principal component analysis; **PDX**, primary tumor derived xenograft; **PR**, progesterone receptor; **PT**, primary tumor; **RT-PCR**, Real time-PCR; **TNBC**, Triple negative breast cancer; **V1**, Visit 1 (before treatment); **V2**, Visit 2 (after treatment).

Article category: Tumor Markers and Signatures

Novelty and Impact

This work describes for the first time the development of a CTCs-derived xenograft (CDX) from a patient with metastatic Triple Negative Breast Cancer, laying on the table the usefulness of this type of approach in the field of liquid biopsy. This study has allowed the identification of prognostic markers and therapeutic targets in CTCs from TNBC patients. The potential clinical translation of the results makes it extremely interesting for oncologists and patients with a high aggressive disease.

Abstract

Triple-negative breast cancer (TNBC) is characterized by high rates of metastasis and no available molecular targets. CTCs derived xenografts (CDX) have demonstrated to be a promising tool for understanding cancer biology. In this study a CDX from a TNBC patient was developed for the first time. After CDX characterization, WNT signalling was found as the main mechanism related with this tumor biology and potential CTCs markers were identified and subsequently validated in TNBC patients. In this cohort high levels of MELK expression were associated with poorer survival rates. Overall, this study demonstrates that CTCs from TNBC are tumorigenic and CDXs are a useful model to obtain valuable information about the tumor.

Introduction

Triple Negative breast cancer (TNBC) is a complex and aggressive subtype of breast cancer (BC) that represents 10-20% of all the cases [1], which is characterized by a lack of both estrogen receptor (ER) and progesterone receptor (PR) expression, and non-amplification of the Human Epidermal growth factor-2 (HER2), which hinders the choice of standard therapy. Depending on its staging TNBC patients are treated with a combination of surgery, chemotherapy and radiation. Heterogeneity of TNBC and the quick development of resistance, greatly limit the clinical benefits of many therapies, explaining their high rate of metastatic disease and their poor overall survival. Consequently, it is essential to delve into the biology of this BC subtype to identify new potential therapeutic targets that can lead to effective treatments. The integration of basic, preclinical and clinical research is the best way to reach it.

In vivo models have a great potential for basic and preclinical cancer research directed to the discovery of biomarkers, the understanding of drug resistance mechanisms and the development of new therapies. In fact, efforts to use PDX (primary tumor derived xenograft) have increased as an *in vivo* model system to study the biology of human tumors since PDX

retain morphologies, architectures and molecular signatures very similar to those of the original tumors [2].

Normally, both tumor molecular characterization and PDX generation are based on tissue biopsies from the primary tumor. However, the opportunity to gain insight into circulating tumoral elements opens a new perspective for tumor characterization. In this sense, it is important to remark that cells dissemination from the primary tumor through the bloodstream as circulating tumor cells (CTCs) is a critical step for the metastasis formation. The enumeration of CTCs has been associated with the clinical outcome in both early and advanced BC patients in a high number of studies [3–8]. CTCs changes, evaluated during treatment, turned to be a reliable surrogate marker of response to treatment. Importantly, CTCs are considered a valuable "liquid biopsy" because they represent the most aggressive tumor population, being a useful model to study the tumor biology at real time, before and during treatment. Despite their potential, there are still technical limitations regarding the molecular characterization and the *ex vivo* expansion of CTCs, mainly due to their scarcity and low viability in circulation.

Transient CTCs primary cultures from metastatic BC patients were obtained[9 10] but, until now, only two groups were able to generate CDX (CTCs-derived xenograft) from BC patients, demonstrating the feasibility to expand these cells *in vivo* and their tumorigenic potential [11,12]. Despite these hopeful results obtained in breast tumors, *in vivo* expansion of CTCs from TNBC patients has not been reached so far.

Here, we describe for the first time the generation of a CDX model from a metastatic TNBC patient, showing the tumorigenic capability of the CTCs population in this patient. CDX molecular characterization provides a unique tool to enhance our understanding of the main pathways implicated in TNBC evolution and dissemination, and to identify interesting prognosis markers and therapeutic targets.

Materials and Methods.

Patients

A total of 32 patients diagnosed of TNBC at Complexo Hospitalario Universitario of Santiago de Compostela were included in the study (Supplementary Table I). In addition, 22 age matched healthy controls were also included. All participants signed an informed consent specifically approved for this study by the Galician Investigation Ethical Committee (code of approval: 2013/462).

Clinical samples

Two tubes (7.5 mL) of peripheral blood were obtained from each patient: one EDTA vacutainer (Becton Dickinson) for CTCs enrichment and characterization, and one CellSave Preservative tube (Menarini, Silicon Biosystems Inc), for CellSearch enumeration. In patients with clear progressing disease, we obtained a second sample, visit 2 (V2).

In patient #20, a 43 years old woman with high CTC counting (969 CTCs/7,5mL) at V2, an additional EDTA blood tube (7.5 mL) was obtained for CDX generation (Figure S2). Formalin-fixed paraffin-embedded (FFPE) biopsies from this patient's primary tumor (obtained at baseline, before surgery and chemotherapy) and metastasis after disease progression were provided by the Pathology Service and the BioBank of Complexo Hospitalario Universitario de Santiago (CHUS) (PT17/0015/0002), integrated in the Spanish National Biobanks Network. These samples were processed following standard operating procedures with the appropriate approval of the Ethical and Scientific Committees.

CTC enumeration and characterization.

CellSave tube was analyzed for CTCs enumeration by the CellSearch System, using CellSearch Epithelial Circulating Tumor Cell Kit (Menarini, Silicon Biosystems Inc); EpCAM enriched cells were labelled with phycoeritrin (PE) conjugated anti-cytokeratins (CKs) antibodies, with

allophycocyanin (APC) conjugated anti-CD45 antibodies and with 4,6-diamino-2-phenylindole (DAPI) to identify the nucleus. The CellTracks Analyzer (Menarini, Silicon Biosystems Inc) was used to acquire digital images of the 3 different fluorescent dyes using a 12-bit camera which were reviewed by trained operators in order to determine the CTCs count.

EDTA tube was used for EpCAM+ isolation of CTCs (CELLection^{TM} Epithelial Enrich Kit, Thermo Fisher Scientific) according to manufacturer's instructions. CTCs coupled to the magnetic beads were resuspended in 100 µl of RNAlater (Ambion) and stored at -80ºC until RNA extraction.

Mice

Mice experimental protocols were approved by the Ethical Committee of the University of Santiago de Compostela (15010/2015/001). Mice were held in our animal facility at the Center for Research in Molecular Medicine and Chronic Diseases (CIMUS, Santiago de Compostela) and given food and water *ad libitum*, in accordance with CIMUS guidelines (ES150780275701). NMRI-Foxn1nu/nu was obtained from Janvier Lab (France) and Scid beige mice (RRID: IMSR_CRL:250) were obtained from the Barcelona Biomedical Research Park (PRBB, Barcelona). After mice arrival, at least one week of acclimation was considered.

CDX establishment & follow-up

For CDX establishment, isolation of peripheral blood mononuclear cell fraction (PBMCs) was performed by density gradient centrifugation protocol (Histopaque®-1077, Sigma) in SepMate™ tubes (StemCell Technologies) according to manufacturer´s instructions. The recovered cells were diluted 1:2 with Matrigel Matrix (Corning) and implanted subcutaneously into a NUDE mouse. After cell injection, the mouse was followed up weekly for tumor development using XenoLight RediJect 2-DG-750 (Perkin Elmer) by Xenogen IVIS 200 system; briefly, 100 µl of the reagent were injected intraperitoneally and the fluorescence was read 3 hours later. Five months after cell injection, a macroscopic tumor was observed; mouse was euthanized due to ethical reasons and the tumor was collected (CDX1). A piece of the tumor

explant (25%) was implanted subcutaneously into a Scid Beige mouse (CDX2), and another piece (25%) was mechanically disaggregated and culture in RPMI media in suspension. After 13 days of culture, cells were collected and injected into the mammary fat pad of a Scid Beige mouse (CDX2M). Tumor growth was monitored by *in vivo* image weekly; CDX2 was euthanized two months after cell injection, and CDX2M after three and a half months. CDX-derived tumor tissue fragments were collected from necropsied animals into RNAlater (Ambion) and stored at −80°C.

CTCs analyses from mice blood

Blood extraction was performed by cardiac puncture using 4 mL EDTA tubes. CTCs enumeration was carried out using CellSearch Technology (Menarini, Silicon Biosystems Inc) from 150 µl of mouse blood mixed with 25 µl of human healthy blood (essential for the correct autofocus of the sample). The volume-scaled protocol for isolation and immunostaining was performed manually using CellSearch Epithelial Circulating Tumor Cell Kit (Menarini, Silicon Biosystems Inc). Digital images acquisition was performed as described before.

RT-qPCR

CTCs

Total RNA from CTCs was extracted with the QIAmp viral RNA mini kit (Qiagen). Reverse transcription was performed with SuperScript III (Invitrogen) following manufacturer's instructions. RT-PCR analysis was performed as described by Barbazán et al.[13] for a customized panel of 19 genes. Expression values for each gene were referred to CD45 as a marker of non-specific isolation.

CDX tumors & FFPE tissue

CDX tumors were disaggregated with Tissuelyser (Qiagen) and RNA extraction performed with AllPrep DNA/RNA Mini Kit (Qiagen) following the manufacturer's instructions. Reverse transcription was performed with SuperScript III (Invitrogen).

Total RNA from FFPE tissue sections was extracted using the miRNeasy FFPE kit (Qiagen). cDNA was synthesized with MulV retrotranscriptase chemistry (Applied Biosystems) following manufacturer's instructions. RT-qPCR analyses were performed in healthy, tumoral and metastatic tissue. Samples were preamplified with TaqMan PreAmp Master Mix (Applied Biosystems) with 14 reaction cycles.

cDNA expression was analyzed on a LightCycler 480 II (Roche Diagnostics) with TaqMan Gene Expression Master Mix and TaqMan probes (Applied Biosystems) for a customized panel of 12 genes in the FFPE tissue and 22 genes in the CDX tumors. GAPDH and B2M were used as reference genes in the FFPE samples and CDX tumors respectively.

Gene expression heatmap was obtained using R software (version 3.4.4.).

Immunohistochemistry

FFPE mouse tissue blocks were sectioned at 3 µm, dried for 1h at 65ºC before pre-treatment procedure of de-paraffinization, rehydration and epitope retrieval in the Pre-Treatment Module, PT-LINK (DAKO) at 95 °C for 20 min in 50x Tris/EDTA buffer, pH 9 and endogenous peroxidase was blocked. The antibodies used recognizes KI67 (1:100, clone SP6), Wide Spectrum Cytokeratin (1:100, Polyclonal), CD45 (1:100, clone EP322Y), ER (1:100, clone SP1), PR (1:100, clone YR85), ECAD (1:100, clone EP700Y), ALDH1A1 (1:100, Polyclonal) (ABCAM, Cambridge, UK); N-Cadherin (1:100, clone EPR1792Y, Merk Millipore, USA), SNAI 1 (1:100, clone H-130), EP-CAM (1:50, clone H70, (Santa Cruz Biotechnology, Dallas, USA), c-erbB-2 (1:100, Polyclonal, Agilent Technologies-DAKO, Santa Clara, United States). After overnight

incubation of the primary antibody, the reaction was visualized with Biotin-SP-AffiniPure Goat Anti-Rabbit IgG (1:200, Jackson ImmunoResearch, Jennersville, USA), and Streptavidin (1:400, Agilent Technologies-DAKO, Santa Clara, United States) using diaminobenzidine chromogen as a substrate. Sections were counterstained with haematoxylin. Appropriate negative controls including no primary antibody were also tested.

The Ki67 percentage was automatically measured using an ACIS® III Instrument (Dako, Denmark, Glostrup). A percentage ≥ 14% was considered as high expression.

RNA-Sequencing Analysis.

Samples were barcoded and prepared for sequencing at the Wellcome Trust Centre for Human Genetics, Oxford, where 75 bp paired-end (PE) reads were obtained on an Illumina HiSeq 4000. The raw data has been deposited in the NCBI's Sequence Read Archive (SRA) under accession number (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA464335). The quality of the sequencing output was assessed using FastQC v.0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Quality filtering and removal of residual adaptor sequences was conducted on read pairs using Trimmomatic v.0.35 [14]. Specifically, Illumina adaptors were clipped from the reads, leading and trailing bases with a Phred score < 20 were removed and the read trimmed if a sliding window average Phred score over four bases was less than 20. Only reads where both PE reads had a length greater than 36 bp post-filtering were retained. Filtered reads were mapped to the human genome (GRCh38.p10) and the mouse genome (GRCm38.p5) using STAR v.2.5.2b [15], the maximum number of mismatches for each read pair was set to 10 % of trimmed read length, and minimum and maximum intron lengths were set to 20 bases and 1 Mb respectively. For those reads aligning to both the human and the mouse genome, only those with a higher mapping quality in human were retained for further analysis.

Paired-reads uniquely mapped to the human genome were counted and assigned to genes using FeatureCounts [16], included in the SourceForge Subread package v.1.5.0. Only reads with both ends mapped to the same gene were retained. Gene count data were used to estimate differential gene expression using the Bioconductor packages DESeq 2 v.3.4 [17]. Samples were hierarchically clustered according to gene read counts after a variance stabilizing transformation, using Euclidean as the distance measure and complete-linkage as the agglomeration method (R package flashClust[18]). Heatmaps of gene expression were created using the R package gplots v3.0.1 heatmap.2 function, using read counts after regularized log transformation (DESeq2[17]).

The Genome Analysis Toolkit (GATK) framework was used for variant calling. Duplicated reads were removed using Picard v.1.128 (http://broadinstitute.github.io/picard). GATK[19] was used to remove sequences overhanging into the intronic regions, to reassign STAR mapping qualities to default values and to perform base quality score recalibration. Somatic SNPs and Indels were identified using Mutect2, which combines the original MuTect[20] and HaplotypeCaller [19]. The mutations were annotated using Variant Effect Predictor (Ensembl version 90[21]).

Enrichment analysis of Gene Ontology terms was done upon uploading selected probe sets identifiers into GSEA (Gene Set Enrichment Analysis) and Panther web tools.

Statistical Analysis.

Data was analyzed using IBM SPSS Statistics for Macintosh, Version 22.0 (SPSS Inc., Chicago) and GraphPad Prism 6.01 software (GraphPad Softwares Inc, San Diego). Gene expression differences between patients and controls were analyzed using Mann-Whitney test. PFS and OS were analyzed using Kaplan-Meier analysis and differences were examined by log-rank test. *p* values < 0.05 were considered statistically significant.

Results

Molecular profiling of CTCs during tumor evolution

One patient with advanced TNBC (Patient #20) was selected for CDX generation because she presented a high CTC count after CellSearch analysis (969 CTCs/7,5mL). A longitudinal study was carried out throughout the disease evolution of this patient (Figure 1 and M&M). Samples were collected at two different time points of the disease to perform CTCs molecular characterization by RT-qPCR which included a panel of genes related with epithelial (EpCAM, ECAD, EGFR, CRIPTO), mesenchymal (VIM, SNAI1) and stem cell (ALDH1, CD49f) features, together with tumor progression associated genes (CD44, BCL11A or AR). CTCs enumeration by CellSearch was 5 CTCs and 969 CTCs/7,5mL in Visit 1 (V1) and Visit 2 (V2), respectively. Figure 2 shows the expression level of the analyzed genes in both CTCs and tumoral tissues (primary tumor and lymph node metastasis). VIM and CD44 showed high expression in all samples while EGFR and AR had low levels. CTCs express higher levels of CRIPTO1 and TIMP1 than tissue samples, while the contrary occurs for SNAI1 expression. CD49f, EpCAM and ECAD increased their expression levels in V2 when comparing with V1. However, when comparing CTCs with tissue samples, EpCAM expression is higher in primary tumor while ECAD expression in V2 is comparable with metastatic samples. In CTCs samples KRT19, TWIST and CTNNB1 were over-expressed in V2 compared with V1 (data not shown).

CTCs-Derived Xenograft development from a TNBC patient is representative of primary tumor

In order to get a TNBC CDX development, the PBMCs fraction (with low percentage of CD45+ cells, data not shown) was injected in an immunocompromised mouse (CellSearch enumeration = 969 CTCs, including 74 CTCs clusters (from 2 to 7 cells) (Figure S1)) and 5 months later mouse was sacrificed and tumor was removed (CDX1) (Figure 3). Part of the tumor was subsequently passaged to 2 Scid Beige mice (CDX2, CDX2M) (see M&M, Figure 3B,D) and monitored over time. In the orthotopic xenograph (CDX2M) blood was collected at

mouse sacrifice and analyzed by CellSearch detecting a CTCs cluster (Figure 3C), which reveals the invasiveness potential of these tumor cells. The pathology analysis disclosed that the three CDX tumors were poorly differentiated carcinoma specimens with high proliferative activity (high ki67 expression: 40% in CDX1, CDx2 and CDX2M). This matched with the primary tumor molecular features: negative for CD45, ER and PR expression and positive for pan-cytokeratines and ECAD (Figure 4A, Figure S2). Expression for EpCAM, N-CAD, ALDH1A1 and SNAI1 was also observed (Figure 4A, Figure S2). Gene expression profile of CDX samples was performed by RT-qPCR to determine similarity among tissue and CTCs samples (Figure 4B) and to check if tumor passaging modifies the genomic profile (Figure S2). Most of the analyzed genes did not change through the passages but it was observed a decrease in AR, ECAD, EGFR and CCND1 expression rate on the second generation mice. An increase in the expression of VIM and a decrease in the expression of CRIPTO1 were observed in CDX2M and CDX2 respectively.

WNT pathway role in tumor progression in the CDX case

RNA sequencing analysis was performed to determine whether CDX samples shared molecular characteristics with patient samples and to identify potential pathways involved in tumor progression in TNBC. For that, healthy or normal tissue (N), primary tumor (PT) and two different metastatic sites (M1 and M2: lymph nodes) from FFPE samples and CDX samples (fresh tissue) were included. First, primary tumor was compared with healthy tissue and, after removing lowly expressed genes (<100 reads), genes showing a $\log_2$ Fold Change |2| were selected as specifically altered due to tumoral transformation (Supplementary Table II). These genes behaviour in all samples was interpreted using a heatmap (Figure 5A), and a principal component analysis (PCA) was carried out to understand how the tumor samples clustered based on their differences with the control (Figure 5B). CDX samples grouped together near the primary tumor. A set of 3401 up-regulated genes and 2372 down-regulated genes were

obtained from this analysis (Supplementary Table II). Venn graph, which shows those genes shared among samples, is represented in Figure 6 and S3. Thus, CDX tissues have 1080 up-regulated genes in common among the three tumors (CDX1, CDX2 and CDX2M) (Figure S3A). M1 and M2 showed 1235 (36.3%) common up-regulated genes (Figure S3B). Then, those common genes were compared between CDXs and metastasis, finding 433 (23%) mutual genes (Figure 6A). We also found 1706 (71.9%) common down-regulated genes in CDX tumors (Figure S3C), while 1079 (45.5%) down-regulated genes were shared between metastatic samples (Figure S3D). Comparative analysis of common genes inferred 823 (41.9%) mutual genes (Figure 6C).

Amongst up-regulated genes in both metastasis and CDX (433), cell cycle genes were strongly represented (Figure 6B), while down-regulated genes (823) accounted for different general cell functions such as system process or tissue development (Figure 6D).

Next, we performed a gene ontology analysis (GO). We included the up-regulated genes from all samples compared to normal tissue. Amongst those genes, WNT signalling was the main pathway involved in the biology of all samples (Figure S4). Further, GO analysis was performed considering the common genes between CDX, metastasis and both. Again, WNT pathway was the most represented followed by cadherin signalling pathway (Figure 7). Regarding down-regulated genes, GnRH, inflammation and WNT pathway stood out (data not shown).

SNPs analysis demonstrated that the three CDXs were very homogeneous since they share 68.4% of the deleterious polymorphisms found (Figure S5A). This group of common SNPs was selected for comparison with the other samples, showing important tumor heterogeneity among them, especially marked between M1 and M2, which only share 1,4% (Figure S5B). While these metastases only share 3 SNPs with the primary tumor, the CDXs samples have 13 SNPs in common with it. There was only one deleterious mutation shared by all the analyzed samples, which was situated in the Cyclin I gene (CCNI, g.77058527A>G) (Figure S5C).

MELK: a prognostic marker for TNBC identified from CDX molecular characterization

A panel of genes was selected for a further validation based on their representation of the main pathways identified by GO analysis (Figure 7 and Figure S4), their expression level in RNA-seq analysis and their involvement in breast cancer development. Genes were analyzed by RT-qPCR in an EpCAM positive CTCs population isolated from the V1 and V2 in patient #20. Gene expression of the 5 selected genes (AURKB, HIST1H4A1, MELK, MYCL and PCDHA8) was detected in both sampling points (data not shown). Lastly, we analyzed these genes in CTCs from the cohort of TNBC patients (n=32, which includes patient #20) and healthy donors (n=22). The AURKB, HIST1H4A1, MELK and PCDHA8 genes were more expressed in patients than in controls ($p < 0.05$) (Figure 8A) demonstrating their presence in the CTC population. In addition, we explored the prognostic potential of our CTCs markers by Kaplan–Meier survival analysis (Supplementary Table III) and we found that MELK overexpression was statistically associated with shorter PFS and OS rates (Figure 8B, C, Supplementary Table III). Besides, although differences were not statistically significant, patients with high levels of AURKB, HIST1H4A1 and PCDHA8 showed also lower survival rates.

Discussion

TNBC subtype is the most challenging among BC tumors due to its aggressive nature, high metastatic potency and lack of targeted therapy. Thus, there is an urgent need to understand the underneath mechanisms involving TN tumors development, which will lead to improvements in the clinical management of these patients. We attempted to tackle this challenge by analysing CTCs in TNBC patients. Importantly, we generated for the first time a CDX from one TNBC patient with high number of CTCs and CTCs clusters. Thus, contrary to a previous work [22], we demonstrated that CTCs from a TNBC patient are tumorigenic and constitute an attractive *in vivo* model to obtain a better understanding of the tumor biology. Histological analysis revealed that the obtained CDX matched with patient tumor phenotype.

Similarly, RNA-Seq analysis also demonstrated that CDX had a common origin with the tumor samples from the patient, confirming that CTCs isolated from blood were tumorigenic, since they were able to reproduce a tumor in a mouse model and subsequent passages. CDX tumors also had metastatic capacity, evidenced by the detection of CTCs in mouse blood. In addition, the detection of CTCs clusters in mouse blood supports the metastatic potential of these cells.

We analyzed different tumor and blood samples over time, from both patient and the CDX, in order to perform a molecular tracking of the disease. Thus, we detected molecular changes among the primary tumor and the metastasis sites as well as among CTCs (V1 and V2), showing the relevance of liquid biopsy monitoring as a valuable tool for understanding tumor evolution. In addition, when CDX and CTCs are compared (V2) there is concordance in the gene expression of BCL11, CD49f, CRIPTO1, ECAD or VIM. The expression of this set of genes, that involved in stem and mesenchymal characteristics, represents a similar phenotype to that observed in patients with TNBC, which despite being an epithelial tumor type has a high mesenchymal phenotype. It is worth mentioning that the expression level of the analyzed genes is not entirely comparable among CTCs and CDX samples as CTCs were obtained using an EpCAM positive enrichment while CDX samples come from a population of CTCs that comprises EpCAM + and EpCAM– CTCs after *in vivo* development.

In addition, the analysis of the CDX allowed us to identify key molecular mechanisms involved in TNBC development that could be important therapeutic target. Thus, GO analysis pointed to the WNT pathway as the main underlying signalling up-regulated in all analyzed samples. WNT has been associated previously with a higher risk of metastasis and worse prognosis in TNBC patients [23,24]. In addition, *in vitro* studies have shown that WNT pathway is preferentially activated in TNBC subtypes, and may represent a possible therapeutic target to treat this type of cancers [25]. Importantly, WNT/β-catenin status identifies patients who are most likely to develop lung and brain metastases[26]. In our data, β-catenin expression is increased in CTCs

isolated in the V2 (more advance disease) compared to V1, and in the three CDX generated from those CTCs, with more than two-fold expression compared with healthy tissue. In addition to the WNT pathway, up-regulated common genes among CDX samples and metastasis samples were mainly involved in cell cycle regulation, in accordance with the high proliferative activity required to form new tumor locations.

Comprehensive data analyses led to the identification of highly expressed genes on all tumor tissue samples, suggesting their important relevance in tumor progression in #20 patient. 5 selected genes were further analyzed in CTCs isolated from this patient and in CTCs from a TNBC patient´s cohort formed by 32 cases, demonstrating that AURKB, HIST1H4A1, MELK and PCDHA8 can be used to detect the presence of CTCs, and therefore, their potential value as indicators of tumor dissemination.

These identified genes have distinct roles in BC development, including TNBC subtype. AURKB is a mitosis-related serine/threonine kinase that is overexpressed in various tumor types including TNBC [27,28]. Elevated AURKB expression contributes to chemoresistance and predicts poor prognosis in BC patients [29], which has led to the development of AURKB inhibitors as anticancer drugs [30,31]. In this regard, the option of monitoring AURKB expression in CTCs from TNBC patients could be an alternative for selecting this type of targeted therapy. Concerning the other analyzed markers, HIST1H4A is a histone cluster member of H4 family and plays a central role for transcription regulation, DNA repair, DNA replication and chromosomal stability. Lai and colleagues verified that acetylation of histone H4 induces cell apoptosis and growth arrest via inhibiting AKT signalling in hepatocellular carcinoma [32], however, little is known regarding its role in other tumor types, including TNBC.

PCDHA8 gene is a member of the protocadherin alpha gene cluster. These neuronal proteins are cadherin-like cell adhesion proteins and have a role in the establishment and functioning of the cell-cell connections that take place in the brain [33]. However, its role *in vivo*, the regulation

of genes or cellular function, have still to be discerned. PCDHA8 is also involved in WNT pathway and has been suggested as a potential marker for the prediction of BC classification and staging of since it was part of a set of genes identified in a computational analysis of tissue-based gene expression data to identify possible gene signatures and markers of blood or urine proteins [34]. We found this gene expressed in CTCs isolated from TNBC patient´s blood samples, however, little is known about its involvement in tumor development or metastasis. Novak and colleagues identified that the PCDHA family preserves aberrant hypermethylation in BC being the overall decrease in the expression of these genes correlated with the increase in the CpG islands methylation of PCDH cluster. This could be due to the disruption of the function of the transcription factors and the regulators of genes involved in the control of their expression[35]. Our results in CTCs support these evidences since we found a reduction in PCDHA8 expression as the disease progressed, probably due to methylation process.

Finally, MELK has been described as an important kinase for the development process and has been implicated in mitotic progression, proliferation, apoptosis, differentiation, stem cell phenotypes, and tumorigenesis [36–39]. This gene has been associated with various types of cancer, particularly aggressive malignancies [40], including TNBC [41–43]. In fact, MELK has been described as one of 22 kinases overexpressed in TNBC when compared with other BC subtypes, and it was also functionally validated *in vitro* [41]. A study using the BC data set of The Cancer Genome Atlas showed that MELK expression was eightfold higher in tumors than in normal breast tissue. In addition, MELK expression correlates with metastatic recurrence and increased mortality indicating that MELK may be predictive of BC metastasis and overall survival rate [44]. Hence, we found that CTCs from TNBC patients express MELK, high levels being associated with lower OS and PFS rates, meaning a difference of 27.25 and 17.7 months, respectively. Regarding CTCs assays, a recent publication described MELK expression in spiked assays using TNBC cell line MDA-MB-231 [45], however, these are not real CTCs. Other work carried out by Fina and colleagues detected MELK expression in CTCs isolated from 7

metastatic breast patients using AdnaTest EMT-1/Stem CellSelect kit [46], reinforcing our results. To our knowledge this is the first time that MELK expression measured in CTCs is correlated with OS or PSF in TNBC patients. Therefore our work marks a milestone pointing MELK as a potential survival marker detected by liquid biopsy, and also a potential therapeutic target with the additional value of already having active MELK inhibitors [47].

Taking into account the interesting results obtained for AURKB, HIST1H4A1, MELK and PCDHA8, further studies should focus on these molecules and their role as tumor markers as well as their implication in tumor biology and tumor dissemination of TNBC subtype.

In conclusion, we integrated CTCs analysis, CDX generation and RNA-Seq technology as a valuable strategy to delve into TNBC biology, providing clinicians with new potential therapeutic targets and markers to improve the clinical management of these patients. We described for the first time a CDX generation from a TNBC patient demonstrating that CTCs from these patients could be tumorigenic in mice. Although CDX generation can not be considered a general approach to improve patient care, it has a great value for translational research, as we have demonstrated. Characterizing the original tumor, the metastasis and the CDX we also confirmed the important role that the WNT pathway plays in TNBC, and identified a panel of markers that can be monitored in CTCs from these patients, providing important information about their tumor aggressiveness and suggesting a possible role in tumor dissemination.

Acknowledgments

We would like to thank the patients for the blood samples; Dr. Fernando Dominguez for providing a nude mouse; Alicia Abalo, Ramón Lago and Carmen Abuín for all the technical support; Dr. de la Fuente and Dra. Alonso for their help with the mice. CiMUS personnel for mice maintenance and CHUS Biobank personnel for their help.

References

1. Cejalvo JM, Martínez de Dueñas E, Galván P, García-Recio S, Burgués Gasión O, Paré L, Antolín S, Martinello R, Blancas I, Adamo B, Guerrero-Zotano Á, Muñoz M, et al. Intrinsic Subtypes and Gene Expression Profiles in Primary and Metastatic Breast Cancer. *Cancer Res [Internet]* 2017;77:2213 LP-2221. Available from: http://cancerres.aacrjournals.org/content/77/9/2213.abstract
2. Cho S-Y, Kang W, Han JY, Min S, Kang J, Lee A, Kwon JY, Lee C, Park H. An Integrative Approach to Precision Cancer Medicine Using Patient-Derived Xenografts. *Mol Cells [Internet]* 2016;39:77–86. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4757806/
3. Wang C, Mu Z, Chervoneva I, Austin L, Ye Z, Rossi G, Palazzo JP, Sun C, Abu-Khalaf M, Myers RE, Zhu Z, Ba Y, et al. Longitudinally collected CTCs and CTC-clusters and clinical outcomes of metastatic breast cancer. *Breast Cancer Res Treat* 2017;
4. Coumans FAW, Ligthart ST, Uhr JW, Terstappen LWMM. Challenges in the enumeration and phenotyping of CTC. *Clin Cancer Res* 2012;
5. Mostert B, Sleijfer S, Foekens JA, Gratama JW. Circulating tumor cells (CTCs): Detection methods and their clinical relevance in breast cancer. *Cancer Treat. Rev.*2009;
6. Giordano A, Cristofanilli M. CTCs in metastatic breast cancer. *Recent Results Cancer Res* 2012;
7. Lianidou ES, Strati A, Markou A. Circulating tumor cells as promising novel biomarkers in solid cancers. *Crit. Rev. Clin. Lab. Sci.*2014;
8. Bidard F-C, Michiels S, Riethdorf S, Mueller V, Esserman LJ, Lucci A, Naume B, Horiguchi J, Gisbert-Criado R, Sleijfer S, Toi M, Garcia-Saenz JA, et al. Circulating Tumor Cells in

Breast Cancer Patients Treated by Neoadjuvant Chemotherapy: A Meta-analysis. *JNCI J Natl Cancer Inst* 2018;

9. Zhang L, Ridgway LD, Wetzel MD, Ngo J, Yin W, Kumar D, Goodman JC, Groves MD, Marchetti D. The identification and characterization of breast cancer CTCs competent for brain metastasis. *Sci Transl Med* 2013;
10. Yu M, Bardia A, Aceto N, Bersani F, Madden MW, Donaldson MC, Desai R, Zhu H, Comaills V, Zheng Z, Wittner BS, Stojanov P, et al. Cancer therapy. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science* 2014;
11. Rossi E, Rugge M, Facchinetti A, Pizzi M, Nardo G, Barbieri V, Manicone M, De Faveri S, Scaini MC, Basso U, Amadori A, Zamarchi R. Retaining the long-survive capacity of Circulating Tumor Cells (CTCs) followed by xeno-transplantation: not only from metastatic cancer of the breast but also of prostate cancer patients. *Oncoscience* 2013;
12. Baccelli I, Schneeweiss A, Riethdorf S, Stenzinger A, Schillert A, Vogel V, Klein C, Saini M, Bäuerle T, Wallwiener M, Holland-Letz T, Höfner T, et al. Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. *Nat Biotechnol [Internet]* 2013;31:539–44. Available from: http://www.nature.com.ez.srv.meduniwien.ac.at/nbt/journal/v31/n6/full/nbt.2576.html%5Cnhttp://www.nature.com/doifinder/10.1038/nbt.2576
13. Barbazán J, Alonso-Alconada L, Muinelo-Romay L, Vieito M, Abalo A, Alonso-Nocelo M, Candamio S, Gallardo E, Fernández B, Abdulkader I, de los Ángeles Casares M, Gómez-Tato A, et al. Molecular Characterization of Circulating Tumor Cells in Human Metastatic Colorectal Cancer. *PLoS One [Internet]* 2012 [cited 2018 Oct 2];7:e40476. Available from: http://dx.plos.org/10.1371/journal.pone.0040476
14. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;

15. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;

16. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;

17. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;

18. Langfelder P, Horvath S. Fast R Functions for Robust Correlations and Hierarchical Clustering. *J Stat Softw [Internet]* 2012;46:i11. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3465711/

19. McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D. b, Gabriel S., Daly M. b, DePristo MA. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;

20. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013;

21. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. *Genome Biol* 2016;

22. Donnenberg VS, Huber A, Basse P, Rubin JP, Donnenberg AD. Neither epithelial nor mesenchymal circulating tumor cells isolated from breast cancer patients are tumorigenic in NOD-scid Il2rgnull mice. *npj Breast Cancer* 2016;

23. DiMeo TA, Anderson K, Phadke P, Feng C, Perou CM, Naber S, Kuperwasser C. A Novel Lung Metastasis Signature Links Wnt Signaling with Cancer Cell Self-Renewal and Epithelial-Mesenchymal Transition in Basal-like Breast Cancer. *Cancer Res [Internet]* 2009;69:5364–73. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2782448/

24. Khramtsov AI, Khramtsova GF, Tretiakova M, Huo D, Olopade OI, Goss KH. Wnt/β-Catenin Pathway Activation Is Enriched in Basal-Like Breast Cancers and Predicts Poor Outcome. *Am J Pathol [Internet]* 2010;176:2911–20. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2877852/

25. Dey N, Barwick BG, Moreno CS, Ordanic-Kodani M, Chen Z, Oprea-Ilies G, Tang W, Catzavelos C, Kerstann KF, Sledge GW, Abramovitz M, Bouzyk M, et al. Wnt signaling in triple negative breast cancer is associated with metastasis. *BMC Cancer* 2013;

26. Bleckmann A, Siam L, Klemm F, Rietkötter E, Wegner C, Kramer F, Beissbarth T, Binder C, Stadelmann C, Pukrop T. Nuclear LEF1/TCF4 correlate with poor prognosis but not with nuclear β-catenin in cerebral metastasis of lung adenocarcinomas. *Clin Exp Metastasis* 2013;

27. Cancer T, Atlas G. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;

28. Marumoto T, Zhang D, Saya H. Aurora-A - A guardian of poles. *Nat. Rev. Cancer*2005;

29. Zhang Y, Jiang C, Li H, Lv F, Li X, Qian X, Fu L, Xu B, Guo X. Elevated Aurora B expression contributes to chemoresistance and poor prognosis in breast cancer. *Int J Clin Exp Pathol* 2015;

30. Kitzen JJEM, de Jonge MJA, Verweij J. Aurora kinase inhibitors. *Crit. Rev. Oncol. Hematol.*2010;

31. Keen N, Taylor S. Aurora-kinase inhibitors as anticancer agents. *Nat. Rev. Cancer*2004;

32. Lai JP, Yu C, Moser CD, Aderca I, Han T, Garvey TD, Murphy LM, Garrity-Park MM, Shridhar V, Adjei AA, Roberts LR. SULF1 Inhibits Tumor Growth and Potentiates the Effects of Histone Deacetylase Inhibitors in Hepatocellular Carcinoma. *Gastroenterology* 2006;

33. Phillips GR, Tanaka H, Frank M, Elste A, Fidler L, Benson DL, Colman DR. Gamma-protocadherins are targeted to subsets of synapses and intracellular organelles in

neurons. *J Neurosci* 2003;

34. Yao F, Zhang C, Du W, Liu C, Xu Y. Identification of gene-expression signatures and protein markers for breast cancer grading and staging. *PLoS One* 2015;
35. Novak P, Jensen T, Oshiro MM, Watts GS, Kim CJ, Futscher BW. Agglomerative epigenetic aberrations are a common event in human breast cancer. *Cancer Res* 2008;
36. Pickard MR, Mourtada-Maarabouni M, Williams GT. Candidate tumour suppressor Fau regulates apoptosis in human cells: An essential role for Bcl-G. *Biochim Biophys Acta - Mol Basis Dis* 2011;
37. Lin ML, Park JH, Nishidate T, Nakamura Y, Katagiri T. Involvement of maternal embryonic leucine zipper kinase (MELK) in mammary carcinogenesis through interaction with Bcl-G, a pro-apoptotic member of the Bcl-2 family. *Breast Cancer Res* 2007;
38. Saito R, Nakauchi H, Watanabe S. Serine/threonine kinase, Melk, regulates proliferation and glial differentiation of retinal progenitor cells. *Cancer Sci* 2012;
39. Joshi K, Banasavadi-Siddegowda Y, Mo X, Kim SH, Mao P, Kig C, Nardini D, Sobol RW, Chow LML, Kornblum HI, Waclaw R, Beullens M, et al. MELK-dependent FOXM1 phosphorylation is essential for proliferation of glioma stem cells. *Stem Cells* 2013;
40. Pitner MK, Taliaferro JM, Dalby KN, Bartholomeusz C. MELK: a potential novel therapeutic target for TNBC and other aggressive malignancies. *Expert Opin. Ther. Targets*2017;
41. Bianchini G, Iwamoto T, Qi Y, Coutant C, Shiang CY, Wang B, Santarpia L, Valero V, Hortobagyi GN, Symmans WF, Gianni L, Pusztai L. Prognostic and therapeutic implications of distinct kinase expression patterns in different subtypes of breast cancer. *Cancer Res* 2010;
42. Speers C, Tsimelzon A, Sexton K, Herrick AM, Gutierrez C, Culhane A, Quackenbush J, Hilsenbeck S, Chang J, Brown P. Identification of novel kinase targets for the treatment

of estrogen receptor-negative breast cancer. *Clin Cancer Res* 2009;

43. Finetti P, Cervera N, Charafe-Jauffret E, Chabannon C, Charpin C, Chaffanet M, Jacquemier J, Viens P, Birnbaum D, Bertucci F. Sixteen-kinase gene expression identifies luminal breast cancers with poor prognosis. *Cancer Res* 2008;

44. Wang Y, Lee YM, Baitsch L, Huang A, Xiang Y, Tong H, Lako A, Von T, Choi C, Lim E, Min J, Li L, et al. MELK is an oncogenic kinase essential for mitotic progression in basal-like breast cancer cells. *Elife* 2014;

45. Porras TB, Kaur P, Ring A, Schechter N, Lang JE. Challenges in using liquid biopsies for gene expression profiling. *Oncotarget* 2018;

46. Fina E, Callari M, Reduzzi C, D'Aiuto F, Mariani G, Generali D, Pierotti MA, Daidone MG, Cappelletti V. Gene expression profiling of circulating tumor cells in breast cancer. *Clin Chem* 2015;

47. Beke L, Kig C, Linders JTM, Boens S, Boeckx A, van Heerde E, Parade M, De Bondt A, Van den Wyngaert I, Bashir T, Ogata S, Meerpoel L, et al. MELK-T1, a small-molecule inhibitor of protein kinase MELK, decreases DNA-damage tolerance in proliferating cancer cells. *Biosci Rep* 2015;

Figure Legends

Figure 1. Patient timeline and clinical case presentation, including the different samples analyzed in our study. M1: Metastasis site 1. M2: Metastasis site 2. V1: Visit 1; V2: Visit 2.

Figure 2. Gene expression analysis. Heatmap depicting RT-PCR expression levels of a panel of genes implicated in TNBC biology from CTCs isolated at Visit 1 (V1, 5 CTCs counted by CellSearch) and Visit 2 (V2, 969 CTCs counted by CellSearch), primary tumor and metastatic tissue. Expression levels were determined based on 33.3 and 66.6 percentiles [Low level (clear

grey) account for 0 to 33.3 percentiles; medium level (dark grey) for 33.3 to 66.6 and high level (black) for 66.6 to 100 percentile].

Figure 3. CDX generation. A) Tumor growth evaluated by *in vivo* imaging 5 months after CTCs injection and tumor development of CDX1. B) Scheme of mammary fat pad distribution in mice. Orthotopic injection of tumor cells disaggregated from CDX1 leads to tumor growth in the mammary fad pat (CDX2M). *In vivo* image of tumor tracking using 2-DG-750. C) CTCs cluster obtained after processing CDX2M mouse blood using CellSearch. D) Macroscopic image of CDX growth after subcutaneous implantation of part of CDX1 tumor (passage 2: CDX2).

Figure 4. CDX Characterization. A) Histological characterization of CDX1 paraffin-embedded samples. Haematoxilin-eosin staining and immunohistochemical analysis of indicated markers (Scale bar: 100 µm). B) Gene expression analysis of CDX1 samples by qRT-PCR. 40-ct normalized by B2M is represented.

Figure 5. RNA sequencing analysis. A) Clustered heat map depicting RNA-normalized expression levels for genes Fold Change |2| to normal sample. B) Principal Components Analysis of indicate samples: normal tissue (N, in black); primary tumor (PT, in blue); metastatic sites (M1 and M2, in red and brown respectively); and CDX1, 2 and 2M (in green).

Figure 6. Venn Diagram analysis. A) Overlapping up-regulated common genes between CDX and metastasis. Grey scale represents percentage of similarity (being darker more similar) B) Gene ontology analysis denoting the main pathways associated to those common up regulated genes (from diagram A). C) Venn Diagrams showing common down regulated genes between CDX and metastasis. D) Gene ontology analysis denoting the main pathways associated to those common down regulated genes (from diagram C).

Figure 7. GO analysis. Gene ontology analysis denoting the main pathways associated to CDX common genes (CDX, dark grey), metastasis common genes (Metastasis, black) and shared genes among CDX and metastasis (CDX+M, in light grey).

Figure 8. Validation of RNA-sequencing results in CTCs from TNBC patients. A) Gene expression levels of AURKB, HIST1H4A1, MELK, MYCL and PCDHA8 genes in a cohort of 32 patients (grey) and 22 controls (white) analyzed by qRT-PCR. *p≤0,05; ** p≤0,01; *** p≤0,001. B and C) Survival curves for progression free survival (PFS) and overall survival (OS) in a cohort of 32 TNBC patients. p was calculated using log-rank test.